BISS Biodiversity Information Science and Standards

Biodiversity Information Science and Standards 2: e26235
doi: 10.3897/biss.2.26235

OPEN ACCESS

Conference Abstract

Modelling Biodiversity Linked Data: Pragmatism May Narrow Future Opportunities

Franck Michel‡, Catherine Faron-Zucker‡, Sandrine Tercerie§, Gargominy Olivier§

‡ Université Côte d'Azur, CNRS, Inria, I3S, Sophia-Antipolis, France
§ Muséum national d'Histoire naturelle, Paris, France

Corresponding author: Franck Michel (franck.michel@cnrs.fr), Gargominy Olivier (gargo@mnhn.fr)

Received: 27 Apr 2018 | Published: 22 May 2018

Citation: Michel F, Faron-Zucker C, Tercerie S, Olivier G (2018) Modelling Biodiversity Linked Data: Pragmatism May Narrow Future Opportunities. Biodiversity Information Science and Standards 2: e26235.
https://doi.org/10.3897/biss.2.26235

Abstract

As the biodiversity community increasingly adopts Semantic Web (SW) standards to represent taxonomic registers, trait banks or museum collections, some questions come up relentlessly: How to model the data? For what goals? Can the same model fulfill different goals?

So far, the community has mostly considered the SW standards through their most salient manifestation: the Web of Linked Data (Heath and Bizer 2011). Indeed, the 5-star Linked Data principles are geared towards the building of a large, distributed knowledge graph that may successfully fulfill biodiversity's need for interoperability and data integration. However, the SW addresses a much broader set of problems involving automatic reasoning. For instance, reasoners can exploit ontological knowledge to improve query answering, leverage class definitions to infer class subsumption relationships, or classify individuals *i.e.* compute instance relationships between individuals and classes by applying reasoning techniques on class definitions and instance descriptions (Shearer et al. 2008).

Whether a "thing" should be modelled as a class or a class instance has been debated at length in the SW community, and the answer is often a matter of perspective. In the context of taxonomic registers for example, the NCBI Organismal Classification (Federhen 2012) and Vertebrate Taxonomy Ontology (Midford et al. 2013) represent taxa as classes in the O

© Michel F et al. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ntology Web Language (OWL). By contrast, other initiatives represent taxa as instances of various classes, *e.g.* the SKOS Concept class (*skos:Concept*) in the AGROVOC thesaurus (Caracciolo et al. 2013) (we speak of the instances as *SKOS concepts*), the Darwin Core taxon class (*dwc:Taxon)* in Encyclopedia of Life (Parr et al. 2016), or classes depicting taxonomic ranks in GeoSpecies, DBpedia and the BBC Wildlife Ontology. Such modelling discrepancies impede linking congruent taxa throughout taxonomic registers. Indeed, one can state the equivalence between two classes (with *owl:equivalentClass*) or two class instances (with *owl:sameAs, skos:exactMatch*, etc.), but good practices discourage the alignment of classes with class instances (Baader et al. 2003).

Recently, Darwin Core's popularity has fostered the modeling of taxa as instances of class *dwc:Taxon* (Senderov et al. 2018, Parr et al. 2016). In this context, pragmatism may incline a Linked Data provider to comply with this majority trend to ensure maximum interlinking. Although technically and conceptually valid, this choice entails certain drawbacks. First, considering a taxon only as a an instance misses the fact that it is a set of biological individuals with common characteristics. An OWL class exactly captures this semantics through the set of necessary and sufficient conditions that an individual must meet to be a class member. In turn, an OWL reasoner can leverage this knowledge to perform query answering, compute subsumption or instance relationships. By contrast, taxa depicted by class instances are not *defined* but *described* by stating their properties. Hence the second drawback: unless we develop bespoke reasoners, there is not much a standard OWL reasoner can deduce from instances.

Yet, some works have demonstrated the effectiveness of logic representation and reasoning capabilities, *e.g.* computing the alignments of two primate classifications (Franz et al. 2016) using generic reasoners that nevertheless require proprietary input formats. OWL reasoners are typically designed to solve such classification problems. They may leverage taxonomic ontologies to compute alignments with other ontologies or apply reasoning to individuals' properties to infer their species. Hence, pragmatically following the instance-based approach may indeed maximize interlinking in the short term, but bears the risk of denying ourselves potentially desirable use cases in the longer term. We believe that developing class-based ontologies for biodiversity should help leverage the SW's extensive theoretical and practical works to tackle a variety of use cases that so far have been addressed with bespoke solutions.

Keywords

Linked Data, modelling, taxonomy, OWL, SKOS, reasoning

Presenting author

Franck Michel, Olivier Gargominy

References

- Baader F, Calvanese D, McGuinness D, Nardi D, Patel-Schneider P (2003) The Description Logic Handbook: Theory, Implementation, and Applications. Cambridge University Press, New York.
- Caracciolo C, Stellato A, Morshed A, Johannsen G, Rajbhandari S, Jaques Y, Keizer J (2013) The AGROVOC linked dataset. Semantic Web 4 (3): 341-348.
- Federhen S (2012) The NCBI Taxonomy database. Nucleic Acids Research 40 (D1): D136-D143-D136-D143.
- Franz N, Pier N, Reeder D, Chen M, Yu S, Kianmajd P, Bowers S, Lud{ B (2016) Two Influential Primate Classifications Logically Aligned. Systematic Biology 65 (4): 561-582. https://doi.org/10.1093/sysbio/syw023
- Heath T, Bizer C (2011) Linked Data: Evolving the Web into a Global Data Space. 1st. Morgan & Claypool
- Midford P, Dececchi TA, Balhoff J, Dahdul W, Ibrahim N, Lapp H, Lundberg J, Mabee P, Sereno P, Westerfield M, Blackburn DC (2013) The Vertebrate Taxonomy Ontology: a framework for reasoning across model organism and species phenotypes. Biomedical Semantics 4 (1): 34-34. https://doi.org/10.1186/2041-1480-4-34
- Parr C, Schulz K, Hammock J, Wilson N, Leary P, Rice J, Corrigan Jr R (2016) TraitBank: Practical semantics for organism attribute data. Semantic Web 7 (6): 577-588. https://doi.org/10.3233/SW-150190
- Senderov V, Simov K, Franz N, Stoev P, Catapano T, Agosti D, Sautter G, Morris R, Penev L (2018) OpenBiodiv-O: ontology of the OpenBiodiv knowledge management system. Biomedical Semantics 9 (1): .
- Shearer R, Motik B, Horrocks I (2008) HermiT: A Highly-Efficient OWL Reasoner. Proc. of the 5th OWLED Workshop.